Weighting the edge stabilization

Alexandre Ern, Jean-Luc Guermond

WEIGHTING THE EDGE STABILIZATION*

ALEXANDRE ERN† AND JEAN-LUC GUERMOND‡§

Abstract. A modification of the edge stabilization technique is proposed to improve the behavior of the method when solving conservation equations with non-smooth data and/or non-smooth solutions. The key ingredient is tempering the edge stabilization in regions of large gradients through appropriate weights. The new method is shown to preserve the convergence properties of the original method on smooth solutions and numerical tests indicate that it performs better on non-smooth solutions.

Key words. Finite elements, conservation equations, edge stabilization, linear stabilization, nonlinear viscosity.

AMS subject classifications. 65M12, 35L65, 65M60

1. Introduction. Linear stabilization techniques are known to be very effective for solving linear first-order PDEs like the transport equation. In particular, denoting h the meshsize and k the polynomial degree of the approximation, linear stabilization techniques yield the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2-norm for smooth solutions. The situation is not so bright when it comes to solving linear problems with non-smooth data and nonlinear conservations equations with non-unique weak solutions. Linear stabilization methods generally promote the Gibbs phenomenon and introduce high-order dissipation that, in the case of nonlinear conservation equations, can lead to convergence to non-entropic solutions.

We focus our attention on the edge stabilization technique [4, 9] as a prototype of linear stabilization, that is relatively easy to implement with H^1-conforming finite elements. We show through numerical examples that edge stabilization promotes the Gibbs phenomenon, and, for nonlinear conservation equations with non-convex flux, cannot select the proper entropic solution. We also show that, when not properly scaled, the extra dissipation induced by edge stabilization can transform a convergent method into a non-convergent one. The purpose of this paper is then to introduce and analyze a modified version of edge stabilization that does not suffer from the above problems. The main modification is tempering, through appropriate weights, the edge stabilization in regions where the discrete solution exhibits large gradients. This may seem a bit counter-intuitive at first glance, since the use of linear stabilization techniques is often motivated to counter spurious oscillations that are produced by large gradients. The proposed method is proved to deliver the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2-norm for smooth solutions. Numerical tests on the linear transport equation in one and two space dimensions show that the weighted edge stabilization performs as required when combined with a nonlinear viscosity method: it no longer promotes the Gibbs phenomenon and does not prevent the nonlinear viscosity method to converge to the correct entropic solution. In other words, the weighted edge stabilization does not antagonize the nonlinear viscosity method. Quite importantly, when combined with a nonlinear viscosity method and for polynomial orders larger than or equal to two, we observe that the weighted edge stabilization increases the convergence order of the nonlinear method in the regions where the solution is smooth. Thus, when combining the weighted edge stabilization with a nonlinear viscosity method, one improves the convergence order of the nonlinear viscosity method without sacrificing its weakened maximum principle property and its ability to properly converge to entropic solutions.

The paper is organized as follows. In §2, we set the notation and present numerical experiments illustrating the main difficulties that are addressed herein. In §3, we introduce and analyze the

*This material is based upon work supported by the National Science Foundation grant DMS-0510650 and by the GNR MOMAS (CNRS/PACEN, ANDRA, BRGM, CEA, EDF, IRSN). Draft version, February 27, 2012

†Université Paris-Est, CERMICS, Ecole des Ponts ParisTech, 77455 Marne-la-Vallée cedex 2, France.

‡Department of Mathematics, Texas A&M University 3368 TAMU, College Station, TX 77843, USA

§On leave from CNRS, France.

weighted edge stabilization method. In §4, we present one- and two-dimensional tests to illustrate the improvements achieved by weighting the edge stabilization. Finally, we draw some conclusions in §5.

2. Preliminaries. The objective of this section is twofold: (i) to set the notation and the model problems we are interested in; (ii) to present numerical experiments that identify the main difficulties that we want address in the present work.

2.1. Formulation of the problem. We are interested in approximating the solution of scalar-valued conservation equations in the form

$$\partial_t u + \nabla\cdot\boldsymbol{f}(u) = 0, \qquad u(x,0) = u_0(x), \quad (x,t) \in \Omega\times\mathbb{R}_+ \tag{2.1}$$

where Ω is an open polyhedral domain in $\mathbb{R}^d$ and $\boldsymbol{f} \in \mathcal{C}^1(\mathbb{R};\mathbb{R}^d)$. For the sake of simplicity, we assume that there are no issues with the boundary conditions; for instance, either we assume periodic boundary conditions or the initial data is compactly supported and we are interested in the solution before the domain of dependence of u_0 reaches the boundary of Ω. We assume that (2.1) has a unique entropic solution satisfying additional entropy inequalities that we omit for brevity.

In order to approximate the entropic solution of (2.1) with H^1-conforming finite elements, we consider a mesh family $\{\mathcal{K}_h\}_{h>0}$ that we assume to be conforming (no hanging nodes) and shape-regular in the sense of Ciarlet. By convention, the elements in $\{\mathcal{K}_h\}_{h>0}$ are closed in $\mathbb{R}^d$. The reference element is denoted $\widehat{K}$ and the map between $\widehat{K}$ and an arbitrary element $K \in \mathcal{K}_h$ is denoted $\Phi_K : \widehat{K} \longrightarrow K$. We define the scalar-valued finite element approximation space

$$X_h = \{v \in \mathcal{C}^0(\Omega;\mathbb{R});\, v|_K\circ\Phi_K \in \mathbb{P}_k,\ \forall K \in \mathcal{K}_h\}, \tag{2.2}$$

where $k \in \mathbb{N}\backslash\{0\}$ and $\mathbb{P}_k$ denotes the set of multivariate polynomials of total degree at most k. For all $K \in \mathcal{K}_h$, h_K denotes the diameter of K divided by k and $h := \max_{K\in\mathcal{K}_h} h_K$ is the so-called meshsize.

Let $t_{\rm F} > 0$ be the final simulation time. We call Galerkin solution of (2.1) the function $u_h \in \mathcal{C}^1([0,t_{\rm F}];X_h)$ such that

$$\int_\Omega v\partial_t u_h \,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot\boldsymbol{f}(u_h)\,\mathrm{d}\Omega = 0, \tag{2.3}$$

for all $v \in X_h$ and all $t \in (0,t_{\rm F})$, and $u_h(t=0) = u_{0,h}$, where $u_{0,h}$ is an appropriate approximation of u_0 in X_h. In all the numerical tests reported herein, the time stepping is done either with the so-called SSP RK3 method, see e.g., [15], or the standard RK4 method. To avoid mixing the space and time discretization errors, we always perform our tests with a small CFL, say CFL = 0.2 or even less. We finally emphasize that the mass matrix is never lumped in our simulations.

The Galerkin solution is known to be a poor approximation of the solution to (2.1) even when the flux is linear. One simple device to stabilize the approximation is to add first-order viscous dissipation. Henceforth, we call viscous solution the function $u_h \in \mathcal{C}^1([0,t_{\rm F}];X_h)$ such that $u_h(t=0) = u_{0,h}$ and

$$\int_\Omega v\partial_t u_h \,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot\boldsymbol{f}(u_h)\,\mathrm{d}\Omega + n_{\rm visc}(u_h;v) = 0, \tag{2.4}$$

for all $v \in X_h$ and all $t \in (0,t_{\rm F})$, with

$$n_{\rm visc}(w;v) := c_{\max}\sum_{K\in\mathcal{K}_h} h_K\|\boldsymbol{f}'(w)\|_{L^\infty(K)}\int_K \nabla w\cdot\nabla v\,\mathrm{d}K. \tag{2.5}$$

We take $c_{\max} = \frac{1}{2k}$ in one space dimension and $c_{\max} = \frac{1}{4k}$ in two space dimensions on triangular meshes.

It is well-known that the viscous solution is only first-order accurate. The performance of the method can be greatly improved by substituting the first-order viscous dissipation by some linear and/or nonlinear stabilization mechanism. We are going to use in this paper the so-called entropy viscosity [20, 18, 21] as a nonlinear stabilization mechanism. We call entropy viscosity solution the function $u_h \in \mathcal{C}^1([0,t_{\rm F}];X_h)$ such that $u_h(t=0)=u_{0,h}$ and

$$\int_\Omega v\partial_t u_h \,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot\boldsymbol{f}(u_h)\,\mathrm{d}\Omega + n_{\rm ev}(u_h;v) = 0, \tag{2.6}$$

for all $v \in X_h$ and all $t \in (0,t_{\rm F})$, with

$$n_{\rm ev}(w;v) := \sum_{K\in\mathcal{K}_h} \nu_K(w)\int_K \nabla w\cdot\nabla v\,\mathrm{d}K, \tag{2.7}$$

where $\nu_K(\cdot) : \mathcal{C}^1([0,t_{\rm F}];X_h) \longrightarrow \mathbb{P}_0(K)$ is a nonlinear viscosity functional. The details on how the nonlinear viscosity is constructed are reported in the Appendix. At this point it is not important to go through the detailed construction of $n_{\rm ev}(\cdot;\cdot)$; it suffices to know that the entropy viscosity solution has reasonable convergence properties. For instance, the entropy viscosity solution has been observed to satisfy a weakened maximum principle in the sense that, for all $\epsilon > 0$, there is $h_0 > 0$ and there are uniform constants c and $\alpha > 0$ such that

$$\|u_h(t)\|_{L^\infty(\Omega)} \le \|u_0\|_{L^\infty(\Omega)} + ch^\alpha, \qquad \forall h < h_0,\ \forall t > \epsilon. \tag{2.8}$$

This property is illustrated, for instance, in Table 2.1(a) below.

When $\boldsymbol{f}(u) = \boldsymbol{\beta}u$, with $\mathbb{R}^d$-valued velocity field $\boldsymbol{\beta}$, (2.1) reduces to the linear transport equation. Many linear (symmetric) stabilization techniques can be considered for solving the linear transport equation, e.g.,, subgrid viscosity [16], edge stabilization [4, 9, 8], and discontinuous Galerkin [24, 14, 22, 12]. All these methods are known to yield the near optimal convergence rate $\mathcal{O}(h^{k+\frac{1}{2}})$ in the L^2-norm for smooth solutions. We focus herein on the edge stabilization technique, which we think is relatively simple to implement with H^1-conforming finite elements. Let $\mathcal{F}_h^i$ be the set of all mesh interfaces. By convention, interfaces are closed in $\mathbb{R}^{d-1}$. For all $F \in \mathcal{F}_h^i$, we denote h_F the diameter of F. Denoting $K_{1,F} \in \mathcal{K}_h$, $K_{2,F} \in \mathcal{K}_h$ the two (closed) cells so that $F = K_{1,F} \cap K_{2,F}$, we set $\Delta_F = K_{1,F} \cup K_{2,F}$, see Figure 3.1 below. Moreover, letting v be a scalar-valued function defined over Δ_F and continuous over $K_{1,F}$ and $K_{2,F}$, we denote $\{v\}(x) = \frac{1}{2}(v|_{K_{1,F}}(x) + v|_{K_{2,F}}(x))$ the average of v at F. The edge stabilization bilinear form is defined as follows:

$$n_{\rm ed}(w;v) = c_{\rm ed}\sum_{F\in\mathcal{F}_h^i} h_F^2\|\boldsymbol{f}'(w)\|_{L^\infty(F)}\int_F \{\partial_n w\}\{\partial_n v\}\,\mathrm{d}F, \tag{2.9}$$

where ∂_n is the outward normal derivative (so that $\{\partial_n w\}$ is proportional to the jump of the normal gradient across F) and $c_{\rm ed}$ is a user-defined parameter. In our numerical experiments, we set $c_{\rm ed} = 0.05$. We call edge stabilized solution the function $u_h \in \mathcal{C}^1([0,t_{\rm F}];X_h)$ such that $u_h(t=0)=u_{0,h}$ and

$$\int_\Omega v\partial_t u_h \,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot\boldsymbol{f}(u_h)\,\mathrm{d}\Omega + n_{\rm ed}(u_h;v) = 0. \tag{2.10}$$

for all $v \in X_h$ and all $t \in (0,t_{\rm F})$.

The objective of this paper is to investigate some aspects of the stabilization properties of the bilinear form $n_{\rm ed}(\cdot;\cdot)$. We want to show that, although this stabilization technique performs

extremely well in smooth regions, it has counter-productive effects in regions of shocks and large gradients. The purpose of the rest of this section is to illustrate through numerical experiments some of the negative effects of $n_{\rm ed}(\cdot;\cdot)$. A weighting technique that cures these problems is proposed and analyzed in §3.

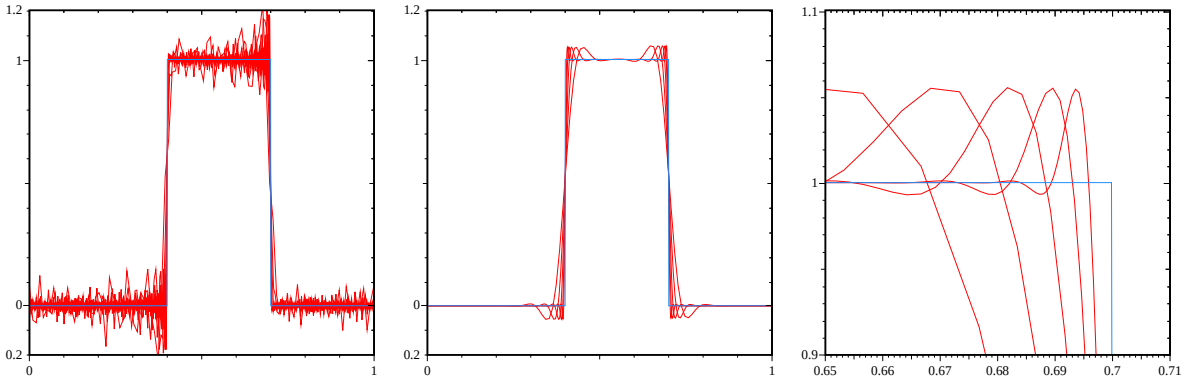
2.2. One-dimensional transport. We consider the one-dimensional transport problem

$$\partial_t u + \partial_x u = 0, \qquad u(x,0) = \begin{cases} 1, & \text{if } 0.4 < x < 0.7, \\ 0, & \text{otherwise,} \end{cases} \tag{2.11}$$

over the interval $\Omega = (0,1)$ with periodic boundary conditions. We compute the solution at $t_{\rm F} = 1$ with continuous $\mathbb{P}_1$ finite elements on various uniform meshes. To assess departures from the maximum principle, we compute the following indicators:

$$e_{\rm Max} := \max_{x\in\Omega} u_h(x,1) - 1, \qquad e_{\rm Min} := -\min_{x\in\Omega} u_h(x,1). \tag{2.12}$$

2.2.1. Edge stabilization alone. We first compute the Galerkin solution on five uniform meshes composed of 100, 2×100, ..., 2^4×100 cells and we set CFL = 0.2. The graphs of the solutions are shown in Figure 2.1(a). We observe the familiar spurious oscillations that characterize



(a) Galerkin solution (b) Edge stabilized sol. (c) Edge stabilized sol. (Zoom)

FIGURE 2.1. *One-dimensional linear transport. Results at* $t_{\rm F} = 1$ *without and with edge stabilization on five uniform meshes composed of* 100, 2×100, ..., 2^4×100 *cells.*

the Galerkin technique. The graphs of the solutions computed with edge stabilization are shown in Figure 2.1(b). This test exemplifies at the same time the stabilizing capability of the edge stabilization and its inability to counter the so-called Gibbs phenomenon triggered in the vicinity of discontinuities and large gradients. Figure 2.1(c) displays details of the graphs of the solutions in the region $x \in [0.65, 0.71]$. Numerical tests on eight refinement levels (nor reported here) show that both indicators $e_{\rm Max}$ and $e_{\rm Min}$ are bounded away from 0, i.e., mesh refinement does not to help satisfy the maximum principle.

2.2.2. Edge stabilization plus entropy viscosity. It is frequently advocated in the literature that linear stabilization must be supplemented with a shock capturing technique to handle properly shocks and large gradients [23, 25, 10, 6, 19]. We want to investigate the effects of combining the edge stabilization with the entropy viscosity described in the Appendix.

We perform the following tests. We compute the entropy viscosity solution and the edge-stabilized entropy viscosity solution of the one-dimensional transport equation (2.11) using $\mathbb{P}_1$ finite

TABLE 2.1
One-dimensional linear transport. Maximum and minimum of entropy viscosity solution (left) and entropy viscosity + edge stabilized solution (right); $\mathbb{P}_1$ finite elements, $t_F = 1$.

(a) Entropy viscosity

h	e_{Min}	rate	e_{Max}	rate
2.500E-03	6.725E-03	–	6.715E-03	–
1.250E-03	5.441E-03	0.306	5.434E-03	0.305
6.250E-04	2.855E-03	0.930	2.854E-03	0.929
3.125E-04	2.235E-03	0.353	2.235E-03	0.353
1.563E-04	1.785E-03	0.324	1.785E-03	0.324

(b) Entropy viscosity + edge stabilization

e_{Min}	rate	e_{Max}	rate
1.597E-02	–	1.597E-02	–
1.600E-02	-0.003	1.600E-02	-0.003
1.633E-02	-0.030	1.633E-02	-0.030
1.626E-02	0.006	1.626E-02	0.006
1.646E-02	-0.017	1.646E-02	-0.017

elements on various uniform meshes, and we evaluate the indicators e_{Max} and e_{Min} defined by (2.12). The results for the entropy viscosity solution and the edge-stabilized entropy viscosity solution are reported in Table 2.1(a) and Table 2.1(b), respectively. We observe that both indicators e_{Max} and e_{Min} for the entropy viscosity solution converge to zero with the meshsize, in agreement with the claim made in Section 2.1 that the entropy viscosity solution satisfies a weakened maximum principle in the form (2.8). On the other hand, we observe in Table 2.1(b) that the weakened maximum principle is lost when edge stabilization is added to the entropy viscosity. Thus, by adding edge stabilization to a method that satisfies a weakened maximum principle, we obtain a method that does not satisfy the maximum principle, even in the weak sense defined above.

The above problem can be easily fixed by increasing the strength of the entropy viscosity. For instance, the weakened maximum principle can be recovered by using $\gamma = \frac{1}{2}$ in the following definition of the entropy viscosity

$$\nu_K(w) = \min(c_{\max} h_K \|\boldsymbol{f}'(w)\|_{L^\infty(K)}, c_{\text{ev}} h_K^\gamma R_K(w)), \tag{2.13}$$

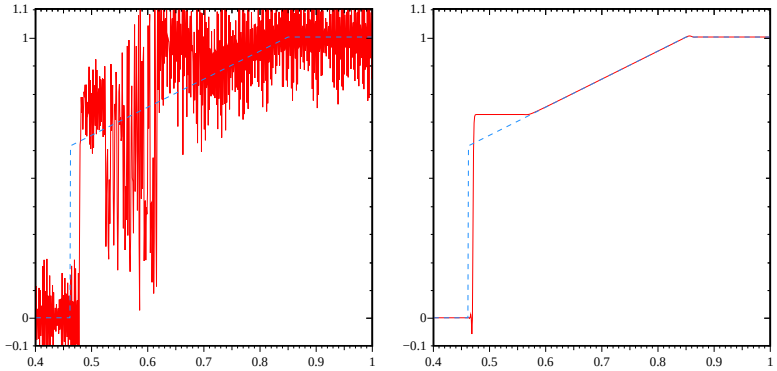
where $R_K(\cdot)$ is the entropy residual (see the Appendix). The definition of ν_K introduced in the Appendix uses $\gamma = 1$, which makes it asymptotically smaller by a factor $h^{\frac{1}{2}}$ than the above definition. The $\gamma = \frac{1}{2}$ fix is marginally satisfactory since it makes the method more dissipative and deteriorates its convergence properties. For instance, convergence tests on the one-dimensional transport problem (2.11) with $\mathbb{P}_1$ finite elements reveal that the convergence rate of the entropy viscosity method in the L^1-norm is $\frac{3}{4}$ with $\gamma = 1$ and $\frac{2}{3}$ with $\gamma = \frac{1}{2}$.

2.3. One-dimensional non-convex conservation equation. We now consider a problem with non-convex flux proposed in [27] to test the edge stabilization technique with nonlinear conservation equations. We restrict ourselves to the one-dimensional domain $\Omega = (0, 1)$ and we consider the following scalar flux and initial data:

$$f(u) = \begin{cases} \frac{1}{4}u(1-u) & \text{if } u < \frac{1}{2}, \\ \frac{1}{2}u(u-1) + \frac{3}{16} & \text{if } \frac{1}{2} \le u, \end{cases} \qquad u_0(x) = \begin{cases} 0, & x \in [0, 0.35], \\ 1, & x \in (0.35, 1]. \end{cases} \tag{2.14}$$

The entropic solution to this problem is a composite wave composed of a shock followed by a rarefaction wave. This problem is challenging since many second-order central schemes with compressive limiters are known to converge to weak solutions that are not entropic. For instance, it is demonstrated in [27] that the so-called central-upwind scheme using second-order piecewise linear reconstruction with either the superbee limiter or the so-called minmod2 limiter fails to converge to the entropic solution. We compute the solution at time $t_F = 1$ with continuous $\mathbb{P}_1$ finite elements. The solution at $t_F = 1$ is composed of a shock wave located at $x_s(1) = \frac{1}{4}(\sqrt{6} - 1)$ followed by a rarefaction wave. The left limit of the solution at the shock is $u_s^-(t_F) = 0$ and the right limit is $u_s^+(t_F) = \frac{2}{x_1(t_F) - x_s(t_F)}(x_1(t_F) - x_0 - \frac{3}{16}t_F) - 1$, where $x_0 = 0.35$ and $x_1(t) = \frac{1}{2}t + x_0$ is the time-dependent location of the head of the rarefaction wave.

2.3.1. Edge stabilization alone. We show in Figure 2.2 the Galerkin and the edge stabilized solutions obtained at $t_{\rm F} = 1$ on a uniform mesh composed of 1000 cells. The time stepping is done with the SSP RK3 scheme, and to avoid time discretization errors, the time step size is based on CFL = 0.01. The solution shown in the left panel of Figure 2.2 is the Galerkin solution and that shown in the right panel is the edge stabilized solution. The entropic solution is shown in blue solid line. It is clear that none of these approximations converge to the entropic solution in any possible norm. The edge stabilized solution is almost free of spurious oscillations but converges to a non-entropic weak solution.



(a) Galerkin solution (b) Edge stabilized sol.

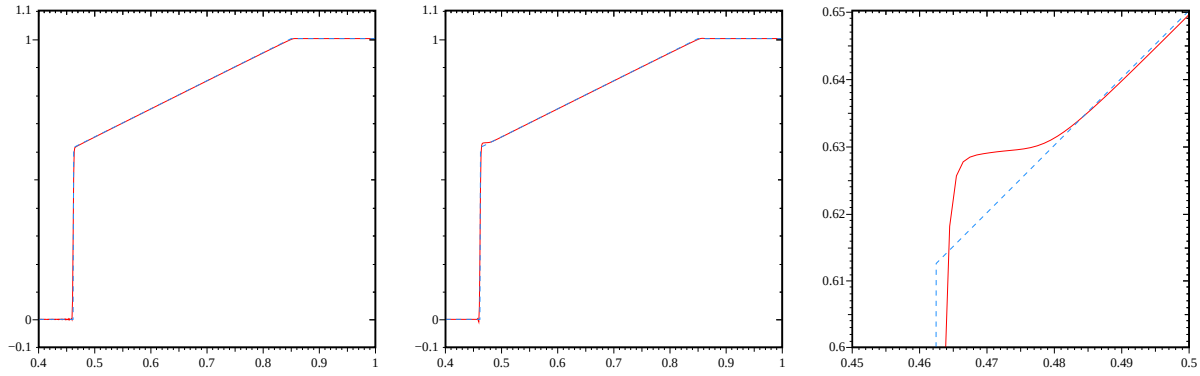
FIGURE 2.2. *Non-convex flux. Results at* $t_{\rm F} = 1$ *on a uniform mesh composed of* 1000 *cells and CFL* $= 0.01$. *Galerkin (left) and edge stabilized (right) solutions.*

2.3.2. Edge stabilization plus entropy viscosity. We show in Figure 2.3 the entropy viscosity solution and the entropy viscosity solution with edge stabilization at $t_{\rm F} = 1$ on a uniform mesh composed of 1000 cells. The entropy viscosity solution is shown in the left panel and the entropy viscosity solution with edge stabilization is shown in the center and right panels. The exact solution (so-called entropic solution) is shown in blue solid line. This tests show that the entropy viscosity solution converges to the entropic solution, whereas the entropy viscosity solution with edge stabilization does not. Thus, by adding edge stabilization to a method that converges to the correct weak solution, we obtain a method that converges to a wrong weak solution. This result is similar to what has been observed in [27] concerning the second-order piecewise linear reconstruction combined with either the superbee or the minmod2 limiter.

2.4. Edge stabilization plus first-order viscosity. In this section, we present two numerical examples showing that the edge stabilization can have adverse effects even on the first-order viscosity method. We first consider the one-dimensional inviscid Burgers equation

$$\partial_t u + \partial_x(\tfrac{1}{2}u^2) = 0, \qquad u(x,0) = \sin(2\pi x), \tag{2.15}$$

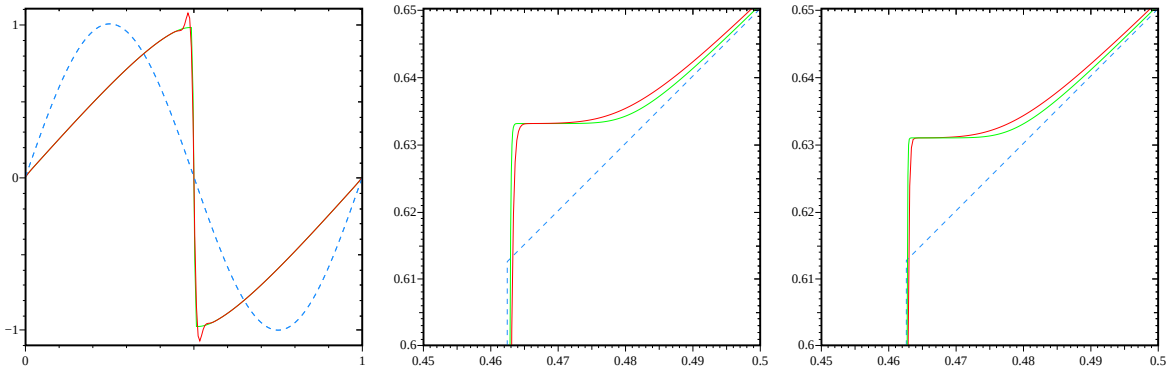
over the interval $\Omega = (0,1)$ with periodic boundary conditions. The solution is computed at $t_{\rm F} = 0.25$ with continuous $\mathbb{P}_1$ finite elements on a mesh composed of 200 elements, and CFL = 0.025. We compute the first-order viscous solution and the first-order viscous solution with edge stabilization. The coefficient $c_{\rm ed}$ in (2.9) is set to 1, and the coefficient $c_{\rm max}$ in (2.5) to $\frac{1}{2}$. When using finite differences on a one-dimensional uniform grid, setting $c_{\rm max} = \frac{1}{2}$ corresponds to replacing the centered differences by first-order upwind differences, and the resulting scheme is known to be monotone. The graph of the two solutions is shown in Figure 2.4(a). For the solution with edge stabilization, we observe over-shoots and under-shoots, and the amplitude of these spurious features is constant as



(a) Entropy viscosity solution (b) Ent. visc. + edge stab. (c) Ent. visc. + edge stab. (zoom)

FIGURE 2.3. *Non-convex flux. Results at* $t_{\rm F} = 1$ *with entropy viscosity without edge stabilization (left) and with edge stabilization on a uniform mesh composed of* 1000 *cells (center and right).*

the mesh is refined. This is characteristic of the Gibbs phenomenon. These over-shoots and under-shoots can be tamed by increasing the viscous dissipation beyond what should normally be necessary. Numerical tests (not shown) reveal that $c_{\max} = 2$ is the lower bound on $c_{\max}$ that makes the viscous dissipation strong enough to overcome the Gibbs phenomenon. These results indicate that edge stabilization (and possibly linear stabilization at large) tends to promote the Gibbs phenomenon.



(a) Burgers equation (b) Non-convex flux, $c_{\rm ed} = 1$ in (2.9) (c) Idem, $c_{\rm ed} = 0.05$ in (2.16)

FIGURE 2.4. *Left: One-dimensional Burgers equation,* $t_{\rm F} = 0.25$*, CFL* $= 0.025$*, uniform mesh composed of* 200 *cells; first-order viscous solution without (in green) and with edge stabilization (in red). Center and right: Non-convex flux,* $t_{\rm F} = 1$*, CFL* $= 0.025$*, uniform meshes composed of* $4,000$ *(in red) and* $10,000$ *cells (in green); first-order viscous solution with edge stabilization (center:* $c_{ed} = 1$ *in* (2.9)*; right:* $c_{ed} = 0.05$ *in* (2.16)*), the entropic solution is shown in dashed blue.*

The adverse effects of edge stabilization are even more dramatic on conservation equations with non-convex flux. We consider again the test case of §2.3. We show in Figure 2.4(b) a zoom of the graph of the solutions obtained with first-order viscosity ($c_{\max} = \frac{1}{2}$) without edge stabilization and with edge stabilization ($c_{\rm ed} = 1$) computed on two uniform meshes composed of $4,000$ and $10,000$ cells. We observe that the viscous solution converges to the entropic solution (as expected), whereas the edge stabilized solution converges to a plateau between the expansion wave and the shock, which

is wrong. The same effect is observed in Figure 2.4(c) with the choice $c_{\rm ed} = 0.05$ in

$$n_{\rm ed}(w;v) = c_{\rm ed}\|\boldsymbol{f}'(w)\|_{L^\infty(\Omega)} \sum_{F\in\mathcal{F}_h^i} h_F^2 \int_F \{\partial_n w\}\,\{\partial_n v\}\,\mathrm{d}F, \tag{2.16}$$

2.5. Conclusions from numerical tests. The first series of tests presented in §2.2.1 shows that the linear edge stabilization does a great job at suppressing spurious oscillations in the regions where the solution is smooth, but this technique cannot get rid of the Gibbs phenomenon. The second series of tests reported in §2.2.2 show that the linear edge stabilization actually promotes the Gibbs phenomenon. The third series of tests on the one-dimensional nonlinear scalar conservation equation with a non-convex flux in §2.3.1 demonstrates that the linear edge stabilization does again a great job at removing the spurious oscillations plaguing the Galerkin solution, but it does not have the correct type of dissipation to make the approximate solution converge to the entropic solution. Finally, the tests reported in §2.3.2 and §2.4 show that not only the linear edge stabilization does not produce the right dissipation, but the type of dissipation that it produces can transform a convergent method (either first-order linear viscosity or nonlinear entropy viscosity) into a non-convergent one. The authors conjecture that the above conclusions are not restricted to edge stabilization but can be extended to some (or most of the) other linear stabilization methods available in the literature.

3. Weighting the edge stabilization. We introduce in this section a weighting technique for the edge stabilization, and we prove that its convergence properties on the linear transport problem are identical to that of the original unweighted edge stabilization in the case of smooth solutions. We consider the linear transport equation

$$\partial_t u + \nabla\cdot(\boldsymbol{\beta} u) = 0, \qquad u(x,0) = u_0(x), \quad (x,t)\in\Omega\times\mathbb{R}_+, \tag{3.1}$$

in space dimension $d = 2$ or $d = 3$. We assume that $\boldsymbol{\beta}$ is Lipschitz and divergence-free in $\overline{\Omega}$. Recall that for simplicity, we assume either periodic boundary conditions, compactly supported solutions, or $\boldsymbol{\beta}\cdot\boldsymbol{n}|_{\partial\Omega} = 0$.

In what follows, $a \lesssim b$ means that inequality $a \le c\,b$ holds with a constant c independent of h (but possibly depending on the mesh-regularity, the polynomial degree k, and the regularity of the problem data and the exact solution). Without loss of generality, we assume $h \le 1$. For any set $R \subset \Omega$ (a mesh element, a mesh face, or a collection thereof), we denote by $\|\cdot\|_{L^p(R)}$ the usual $L^p(R)$-norm, $1 \le p \le \infty$, for scalar- or vector-valued functions.

3.1. Principle of the method. Consider the following discrete solution $u_h \in \mathcal{C}^1([0,t_{\rm F}];X_h)$, with X_h defined in (2.2), so that $u_h(t=0) = u_{0,h}$ and

$$\int_\Omega v\partial_t u_h\,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot(\boldsymbol{\beta}u_h)\,\mathrm{d}\Omega + n_{\rm lim,ed}(u_h;u_h,v) = 0. \tag{3.2}$$

for all $v \in X_h$ and all $t \in (0,t_{\rm F})$. The weighted edge stabilization semi-linear form is defined by

$$n_{\rm lim,ed}(z;w,v) = c_{\rm ed} \sum_{F\in\mathcal{F}_h^i} \alpha(g_F(z))h_F^2|\boldsymbol{\beta}|_F \int_F \{\partial_n w\}\,\{\partial_n v\}\,\mathrm{d}F, \tag{3.3}$$

with the shorthand notation $|\boldsymbol{\beta}|_F := \|\boldsymbol{\beta}\|_{L^\infty(F)}$ and $g_F(z)$ is a measure of the gradient of z around F which we take in the form

$$g_F(z) = \ell^{-1}|\langle\nabla z\rangle_{\Delta_F}|, \tag{3.4}$$

where $\langle\phi\rangle_R := \mathrm{meas}(R)^{-1}\int_R \phi\,\mathrm{d}R$ denotes the average of a function ϕ over a set $R \subset \Omega$ and where the global scaling parameter ℓ is, e.g., set to $\ell := |\langle\nabla u_{0,h}\rangle_\Omega|$. The key ingredient in (3.3) is the function $\alpha : \mathbb{R}_+ \to (0,1]$ which weights the amount of edge stabilization. The function α must be such that

(i) α is non-increasing;
(ii) there is $\alpha_0 \in \mathbb{R}_+$ and $\lambda \in \mathbb{R}_+$ such that, for all $r \geq 1$, $\alpha(r) \geq \alpha_0 r^{-\lambda}$.

Typically, $\alpha(r) \to 0$ as $r \to +\infty$ so as to turn off edge stabilization in regions of large gradients. Condition (ii) then means that α must not decrease too quickly, so as to retain the optimal convergence properties of the method for smooth solutions.

3.2. Convergence analysis. We first prove a convergence result in two dimensions. In this situation, there is no restriction on the parameter λ controlling the decrease of the weighting function α at infinity.

THEOREM 3.1. *Let u and u_h be the solutions to* (3.1) *and to* (3.2)*, respectively. Assume that $u \in C^1(0,t_\mathrm{F};H^{k+1}(\Omega)) \cap C^0(0,t_\mathrm{F};W^{k+1,\infty}(\Omega))$. Assume $d = 2$ and quasi-uniform meshes. Let $\lambda > 0$. Then, for all $t \in [0,t_\mathrm{F}]$,*

$$\|(u-u_h)(\cdot,t)\|_{L^2(\Omega)} + \left(\int_0^t n_{\mathrm{lim,ed}}(u_h;u_h,u_h)\,\mathrm{d}\tau\right)^{\frac{1}{2}} \lesssim h^{k+\frac{1}{2}}. \tag{3.5}$$

Proof. The proof is decomposed into three steps.

Step 1: error equation. For all $t \in [0,t_\mathrm{F}]$, let $w(\cdot,t)$ be the L^2-orthogonal projection of the exact solution $u(\cdot,t)$ onto the discrete space X_h (in the case of compactly supported solutions, we project onto $X_h \cap H_0^1(\Omega)$). We define the quantities (the dependence with respect to t is now left implicit)

$$e := u_h - w, \qquad \eta := u - w,$$

so that the approximation error is $u_h - u = e - \eta$. Observing that

$$\int_\Omega v\partial_t w\,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot(\boldsymbol{\beta}w)\,\mathrm{d}\Omega = -\int_\Omega v\partial_t\eta\,\mathrm{d}\Omega - \int_\Omega v\nabla\cdot(\boldsymbol{\beta}\eta)\,\mathrm{d}\Omega, \qquad \forall v \in X_h,$$

and subtracting this equation from (3.2), we infer, for all $v \in X_h$ and all $t \in [0,t_\mathrm{F}]$,

$$\begin{aligned}\int_\Omega v\partial_t e\,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot(\boldsymbol{\beta}e)\,\mathrm{d}\Omega + n_{\mathrm{lim,ed}}(u_h;e,v) &= \int_\Omega v\partial_t\eta\,\mathrm{d}\Omega + \int_\Omega v\nabla\cdot(\boldsymbol{\beta}\eta)\,\mathrm{d}\Omega + n_{\mathrm{lim,ed}}(u_h;\eta,v)\\ &=: \mathfrak{T}_1(\eta,v) + \mathfrak{T}_2(\eta,v) + \mathfrak{T}_3(u_h;\eta,v),\end{aligned} \tag{3.6}$$

where we have used the fact that, for the exact solution u, $\{\partial_n u\} = 0$ for all $F \in \mathcal{F}_h^i$, so that

$$\begin{aligned}n_{\mathrm{lim,ed}}(u_h;u_h,v) &= n_{\mathrm{lim,ed}}(u_h;e,v) - n_{\mathrm{lim,ed}}(u_h;\eta,v) + n_{\mathrm{lim,ed}}(u_h;u,v)\\ &= n_{\mathrm{lim,ed}}(u_h;e,v) - n_{\mathrm{lim,ed}}(u_h;\eta,v).\end{aligned}$$

Step 2: basic estimates. Testing (3.6) with $v = e$ and using the conservativity property

$$\int_\Omega e\nabla\cdot(\boldsymbol{\beta}e)\,\mathrm{d}\Omega = 0,$$

which holds owing to the choice of boundary conditions and the fact that $\boldsymbol{\beta}$ is divergence-free, we infer that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e\|_{L^2(\Omega)}^2 + n_{\mathrm{lim,ed}}(u_h;e,e) \leq |\mathfrak{T}_1(\eta,e)| + |\mathfrak{T}_2(\eta,e)| + |\mathfrak{T}_3(u_h;\eta,e)|.$$

Moreover, since $u \in C^1(0,t_\mathrm{F};H^{k+1}(\Omega))$, classical finite element interpolation properties [3, 13] yield

$$h_K\|\nabla\eta\|_{L^2(K)} + \|\eta\|_{L^2(K)} + \|\partial_t\eta\|_{L^2(K)} \lesssim h_K^{k+1}, \qquad \forall K \in \mathcal{K}_h. \tag{3.7}$$

Hence, by the Cauchy–Schwarz inequality, we obtain

$$|\mathfrak{T}_1(\eta,e)| \lesssim h^{k+1}\|e\|_{L^2(\Omega)}, \qquad |\mathfrak{T}_2(\eta,e)| \lesssim h^k\|e\|_{L^2(\Omega)},$$
$$|\mathfrak{T}_3(u_h;\eta,e)| \le n_{\mathrm{lim,ed}}(u_h;e,e)^{\frac12} n_{\mathrm{lim,ed}}(u_h;\eta,\eta)^{\frac12} \lesssim h^{k+\frac12} n_{\mathrm{lim,ed}}(u_h;e,e)^{\frac12},$$

where the last bound results from the fact that $n_{\mathrm{lim,ed}}(u_h;\eta,\eta)^{\frac12} \le n_{\mathrm{ed}}(\eta,\eta)^{\frac12}$ (since $\alpha_F(u_h)\le 1$) and the classical bound $n_{\mathrm{ed}}(\eta,\eta)^{\frac12} \lesssim h^{k+\frac12}$. Collecting the above estimates, using Young's inequality, and recalling that $h\le 1$, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}\|e\|_{L^2(\Omega)}^2 + n_{\mathrm{lim,ed}}(u_h;e,e) \lesssim h^k\|e\|_{L^2(\Omega)} + h^{2k+1},$$

whence by using Gronwall's Lemma together with $h\le 1$, we obtain the suboptimal estimate

$$\forall t\in[0,t_{\mathrm{F}}], \qquad \|e(\cdot,t)\|_{L^2(\Omega)} + \left(\int_0^t n_{\mathrm{lim,ed}}(u_h;e,e)\,\mathrm{d}\tau\right)^{\frac12} \lesssim h^k. \tag{3.8}$$

Step 3: improved estimate on $\mathfrak{T}_2(\eta,e)$. Let $\epsilon\ge 0$ and $c_0\ge 1$ (the value of these quantities is chosen later on). We fix a time $t\in(0,t_{\mathrm{F}})$. We define the sets

$$\mathcal{F}_h^\sharp := \{F\in\mathcal{F}_h^i;\ g_F(u_h)\ge c_0h^{-\epsilon}\},$$
$$\mathcal{K}_h^\flat := \{K\in\mathcal{K}_h;\ \forall F\in\mathfrak{F}_K,\ F\notin\mathcal{F}_h^\sharp\},$$
$$\mathcal{K}_h^\sharp := \{K\in\mathcal{K}_h;\ \exists F\in\mathfrak{F}_K,\ F\in\mathcal{F}_h^\sharp\} = \mathcal{K}_h\setminus\mathcal{K}_h^\flat,$$

where $\mathfrak{F}_K$ is the collection of all the mesh interfaces having a nonempty intersection with K (see Figure 3.2).

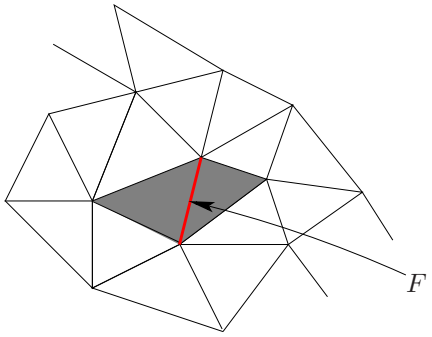


FIGURE 3.1. *Definition of Δ_F (grey triangles)*

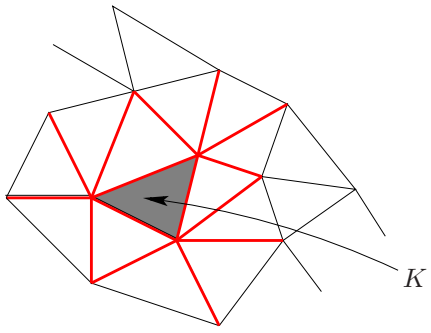


FIGURE 3.2. *Definition of $\mathfrak{F}_K$ (thick lines)*

Let $\boldsymbol{\beta}_h$ be the continuous, piecewise affine interpolant of $\boldsymbol{\beta}$ on the mesh $\mathcal{K}_h$ and set $z_h := \boldsymbol{\beta}_h\cdot\nabla e$. Observe that z_h is a piecewise polynomial of degree $\le k$, but it does not belong to X_h because it can jump across interfaces. We define $\mathcal{I}_{\mathrm{av}}(z_h)\in X_h$ to be so that its value at any Lagrange node is equal to the arithmetic average of the values taken by z_h from the mesh elements sharing that node. It is well-known, see [1, 26, 7], that this type of averaging leads to the estimate, for all $K\in\mathcal{K}_h$,

$$\|z_h - \mathcal{I}_{\mathrm{av}}(z_h)\|_{L^2(K)} \lesssim \sum_{F\in\mathfrak{F}_K} h_F^{\frac12}\|[\![z_h]\!]\|_{L^2(F)} \le \sum_{F\in\mathfrak{F}_K} h_F^{\frac12}|\boldsymbol{\beta}|_F\|\,\{\partial_n e\}\,\|_{L^2(F)}, \tag{3.9}$$

where $[\![\cdot]\!]$ denotes the jump across F and where we have used the continuity of $\boldsymbol{\beta}_h$ across F. Integrating by parts, accounting for boundary conditions, and using the fact that η is L^2-orthogonal to X_h, $\mathfrak{T}_2(\eta,e)$ can be decomposed as

$$\begin{aligned}\mathfrak{T}_2(\eta,e) &= -\int_\Omega \eta(\boldsymbol{\beta}\cdot\nabla e)\,\mathrm{d}\Omega = -\int_\Omega \eta(\boldsymbol{\beta}-\boldsymbol{\beta}_h)\cdot\nabla e\,\mathrm{d}\Omega - \int_\Omega \eta z_h\,\mathrm{d}\Omega\\ &= -\int_\Omega \eta(\boldsymbol{\beta}-\boldsymbol{\beta}_h)\cdot\nabla e\,\mathrm{d}\Omega - \int_{\Omega^\flat} \eta(z_h-\mathcal{I}_{\mathrm{av}}(z_h))\,\mathrm{d}\Omega - \int_{\Omega^\sharp} \eta(z_h-\mathcal{I}_{\mathrm{av}}(z_h))\,\mathrm{d}\Omega\\ &=: \mathfrak{T}_{2,0}(\eta,e)+\mathfrak{T}_{2,1}(\eta,e)+\mathfrak{T}_{2,2}(\eta,e),\end{aligned}$$

where $\Omega^\flat := \cup_{K\in\mathcal{K}_h^\flat} K$ and $\Omega^\sharp := \cup_{K\in\mathcal{K}_h^\sharp} K$. Since $\boldsymbol{\beta}$ is Lipschitz, the term $\mathfrak{T}_{2,0}(\eta,e)$ is readily bounded as follows:

$$|\mathfrak{T}_{2,0}(\eta,e)| \lesssim \|\eta\|_{L^2(\Omega)} h \|\nabla e\|_{L^2(\Omega)} \lesssim \|\eta\|_{L^2(\Omega)} \|e\|_{L^2(\Omega)} \lesssim h^{2k+1},$$

where we have used an inverse inequality and the estimate (3.8) on $\|e\|_{L^2(\Omega)}$. For the term $\mathfrak{T}_{2,1}(\eta,e)$, we use (3.9), the Cauchy–Schwarz inequality, $h_K \lesssim h_F$, and $|\boldsymbol{\beta}|_F \lesssim 1$ to infer

$$|\mathfrak{T}_{2,1}(\eta,e)| \lesssim \left(\sum_{K\in\mathcal{K}_h^\flat} [\alpha(g_K(u_h))h_K]^{-1}\,\|\eta\|_{L^2(K)}^2\right)^{\frac12} \left(\sum_{K\in\mathcal{K}_h^\flat} \alpha(g_K(u_h)) \sum_{F\in\mathfrak{F}_K} h_F^2 |\boldsymbol{\beta}|_F \|\,\{\partial_n e\}\,\|_{L^2(F)}^2\right)^{\frac12},$$

with $g_K(u_h) := \max_{F\in\mathfrak{F}_K} g_F(u_h)$. Since the function α is non-increasing, $\alpha(g_K(u_h)) \le \alpha(g_F(u_h))$ for all $F\in\mathfrak{F}_K$, so that this estimate implies

$$|\mathfrak{T}_{2,1}(\eta,e)| \lesssim \left(\sum_{K\in\mathcal{K}_h^\flat} [\alpha(g_K(u_h))h_K]^{-1}\,\|\eta\|_{L^2(K)}^2\right)^{\frac12} n_{\mathrm{lim,ed}}(u_h;e,e)^{\frac12}.$$

We now use Assumption (ii) on the function α to infer that, for all $K\in\mathcal{K}_h^\flat$ and all $F\in\mathfrak{F}_K$, $\alpha(g_F(u_h)) \ge \alpha(c_0 h^{-\epsilon}) \ge \alpha_0 c_0^{-\lambda} h^{\epsilon\lambda}$ since $c_0 h^{-\epsilon} \ge 1$ (because $c_0\ge 1$, $\epsilon\ge 0$, and $h\le 1$). Thus, we obtain $\alpha(g_K(u_h)) \ge \alpha_0 c_0^{-\lambda} h^{\epsilon\lambda}$, whence

$$\left(\sum_{K\in\mathcal{K}_h^\flat} [\alpha(g_K(u_h))h_K]^{-1}\,\|\eta\|_{L^2(K)}^2\right)^{\frac12} \lesssim h^{k+\frac12-\frac12\lambda\epsilon},$$

so that

$$|\mathfrak{T}_{2,1}(\eta,e)| \lesssim h^{k+\frac12-\frac12\lambda\epsilon} n_{\mathrm{lim,ed}}(u_h;e,e)^{\frac12}.$$

We now bound $\mathfrak{T}_{2,2}(\eta,e)$. Since $\|\eta\|_{L^\infty(\Omega)} \lesssim h^{k+1}$ owing to $u\in W^{k+1,\infty}(\Omega)$ and the approximation properties in $L^\infty(\Omega)$ of the L^2-orthogonal projection for $d=2$ and quasi-uniform meshes, see [11], we infer

$$|\mathfrak{T}_{2,2}(\eta,e)| \le \|\eta\|_{L^\infty(\Omega^\sharp)} \operatorname{meas}(\Omega^\sharp)^{\frac12} \|z_h - \mathcal{I}_{\mathrm{av}}(z_h)\|_{L^2(\Omega^\sharp)} \lesssim h^{2k} \operatorname{meas}(\Omega^\sharp)^{\frac12},$$

where we have used $\|z_h-\mathcal{I}_{\mathrm{av}}(z_h)\|_{L^2(\Omega)} \lesssim \|z_h\|_{L^2(\Omega)} \lesssim \|\nabla e\|_{L^2(\Omega)} \lesssim h^{k-1}$. We now estimate $\operatorname{meas}(\Omega^\sharp)$. This is achieved by proving $\operatorname{meas}(\Omega^\sharp) \lesssim \operatorname{meas}(\Omega^{\Delta\sharp})$ and estimating $\operatorname{meas}(\Omega^{\Delta\sharp})$. Here, $\Omega^{\Delta\sharp} := \cup_{F\in\mathcal{F}_h^\sharp} \Delta_F$, so that $\Omega^{\Delta\sharp}$ is the union of all the elements that have a face in $\mathcal{F}_h^\sharp$. Note that $\Omega^{\Delta\sharp} \subset \Omega^\sharp$; see Figure 3.3.

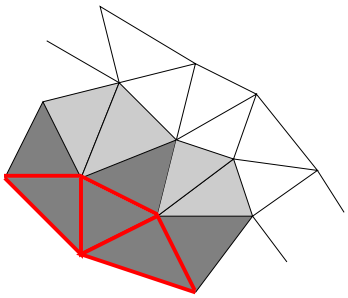


FIGURE 3.3. *The faces in thick red lines are in the set $\mathcal{F}_h^\sharp$; the triangles filled in dark grey are in $\Omega^{\Delta\sharp}$; the triangles filled in dark or light grey are in $\Omega^\sharp$ (and thus belong to $\mathcal{K}_h^\sharp$); unfilled triangles belong to $\mathcal{K}_h^\flat$.*

Consider $K \in \mathcal{K}_h^\sharp$. By definition, there is $F \in \mathfrak{F}_K$ such that $F \in \mathcal{F}_h^\sharp$. By definition also, $\Delta_F \subset \Omega^{\Delta\sharp}$ and $K \cap \Delta_F \neq \emptyset$ (recall that K and Δ_F are closed). This means that there is $K' \in \Omega^{\Delta\sharp}$ such that $K \cap K'$ is nonempty, so that $\mathcal{K}_h^\sharp \subset \cup_{K' \in \Omega^{\Delta\sharp}} \{K'' \in \mathcal{K}_h;\ K' \cap K'' \neq \emptyset\}$. Hence, using mesh regularity, we infer

$$\operatorname{meas}(\Omega^\sharp) \le \sum_{K' \in \Omega^{\Delta\sharp}} \operatorname{meas}(\{K'' \in \mathcal{K}_h;\ K' \cap K'' \neq \emptyset\}) \lesssim \sum_{K' \in \Omega^{\Delta\sharp}} \operatorname{meas}(K') = \operatorname{meas}(\Omega^{\Delta\sharp}).$$

The definition of $\mathcal{F}_h^\sharp$ now implies that

$$\begin{aligned}
\operatorname{meas}(\Omega^{\Delta\sharp}) h^{-2\epsilon} &\le \sum_{F \in \mathcal{F}_h^\sharp} \operatorname{meas}(\Delta_F) h^{-2\epsilon} \le (c_0 \ell)^{-2} \sum_{F \in \mathcal{F}_h^\sharp} \operatorname{meas}(\Delta_F) |\langle \nabla u_h \rangle_{\Delta_F}|^2 \\
&\le (c_0 \ell)^{-2} \sum_{F \in \mathcal{F}_h^\sharp} \int_{\Delta_F} |\nabla u_h|^2 \, \mathrm{d}\Omega \le 2 (c_0 \ell)^{-2} \|\nabla u_h\|_{L^2(\Omega^{\Delta\sharp})}^2 \\
&\le 6 (c_0 \ell)^{-2} (\|\nabla e\|_{L^2(\Omega^{\Delta\sharp})}^2 + \|\nabla \eta\|_{L^2(\Omega^{\Delta\sharp})}^2 + \|\nabla u\|_{L^2(\Omega^{\Delta\sharp})}^2).
\end{aligned}$$

Since $\|\nabla e\|_{L^2(\Omega^{\Delta\sharp})} \lesssim h^{k-1}$, $\|\nabla \eta\|_{L^2(\Omega^{\Delta\sharp})} \lesssim h^k$, and $\|\nabla u\|_{L^2(\Omega^{\Delta\sharp})} \le \operatorname{meas}(\Omega^{\Delta\sharp})^{\frac12} \|\nabla u\|_{L^\infty(\Omega^{\Delta\sharp})}$, choosing $c_0 = \max(\sqrt{12}\ell^{-1} \|\nabla u\|_{L^\infty(\Omega \times (0,t_F))}, 1)$ allows one to hide $\|\nabla u\|_{L^2(\Omega^{\Delta\sharp})}^2$ on the left-hand side, so that $\operatorname{meas}(\Omega^{\Delta\sharp}) \lesssim h^{2(k-1+\epsilon)}$, and hence

$$|\mathfrak{T}_{2,2}(\eta, e)| \lesssim h^{3k-1+\epsilon}.$$

Collecting the above estimates and dropping higher-order terms, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t} \|e\|_{L^2(\Omega)}^2 + n_{\mathrm{lim,ed}}(u_h; e, e) \lesssim h^{3k-1+\epsilon} + h^{2k+1-\lambda\epsilon}.$$

For $k \ge 2$, we can take $\epsilon = 0$ and obtain the desired estimate of order h^{2k+1} since $3k - 1 \ge 2k + 1$ in this case. In the case $k = 1$, the two terms on the right-hand side are balanced by choosing $\epsilon = \frac{1}{1+\lambda}$. This leads to the estimate (3.8) with the sharper bound $h^{1+\rho}$ with $\rho = \frac12 \epsilon$. We now use a bootstrap argument. Suppose that estimate (3.8) holds with bound $h^{1+\rho_n}$. Then, proceeding as above with a parameter ϵ_n to be chosen, the new bound for $\mathfrak{T}_{2,2}(\eta, e)$ becomes $h^{2+\epsilon_n+2\rho_n}$, while the bound for $\mathfrak{T}_{2,1}(\eta, e)$ is still $h^{3-\lambda\epsilon_n}$. Balancing the two bounds leads to the choice $\epsilon_n = \frac{1}{1+\lambda}(1 - 2\rho_n)$, thus improving estimate (3.8) to $h^{1+\rho_{n+1}}$ with $\rho_{n+1} = \frac12(1 - \lambda\epsilon_n) = \frac12(\frac{1}{1+\lambda} + \frac{2\lambda}{1+\lambda}\rho_n)$. This recursive relation shows that ρ_n converges to $\frac12$ (for all λ), thereby completing the proof. □

Remark 3.1. Theorem 3.1 holds in any space dimension under the assumption of locally quasi-uniform meshes provided the parameter $\delta = \sup_{h>0} \max_{K\in\mathcal{K}_h} \max_{K'\cap K\neq\emptyset} |1-(h_{K'}/h_K)^2|$ is small enough, since the estimates $\|\eta\|_{L^\infty(\Omega)} \lesssim h^{k+1}$ and $\|h^{-\frac{1}{2}}\eta\|_{L^2(\Omega)} \lesssim h^{k+\frac{1}{2}}$ hold, see [2].

We now prove a convergence result in three dimensions (with a slightly less stringent regularity assumption on the exact solution). The main difference with the result of Theorem 3.1 is that, at least for the lowest-order polynomials, the convergence result now requires that the parameter λ controlling the decrease of the weighting function α at infinity be not too large.

THEOREM 3.2. *Let u and u_h be the solutions to* (3.1) *and to* (3.2)*, respectively. Assume that $u \in C^1(0,t_\mathrm{F};H^{k+1}(\Omega)) \cap C^0(0,t_\mathrm{F};W^{1,\infty}(\Omega))$. Assume $d=3$ and quasi-uniform meshes. Then, the following holds for all $t\in[0,t_\mathrm{F}]$: if $k\geq 3$ and any λ,*

$$\|(u-u_h)(\cdot,t)\|_{L^2(\Omega)} + \left(\int_0^t n_{\mathrm{lim,ed}}(u_h;u_h,u_h)\,\mathrm{d}\tau\right)^{\frac{1}{2}} \lesssim h^{k+\frac{1}{2}}, \tag{3.10}$$

and the right-hand side becomes $h^{2+\frac{2-\lambda}{4+\lambda}}$ if $k=2$ and $\lambda<2$, and $h^{1+\frac{2-3\lambda}{4+\lambda}}$ if $k=1$ and $\lambda<\frac{2}{3}$.

Proof. We proceed as in the proof of Theorem 3.1, up to the bound on $\mathfrak{T}_{2,2}(\eta,e)$ in step 3 of the previous proof. Here, we no longer use approximation properties in $L^\infty(\Omega)$ of the L^2-orthogonal projection. Instead, owing to the Sobolev embedding, we observe that there holds $\|\eta\|_{L^p(\Omega)} \lesssim h^k$ with $p=6$. Then,

$$|\mathfrak{T}_{2,2}(\eta,e)| \lesssim h^{2k-1}\,\mathrm{meas}(\Omega^\sharp)^{\frac{p-2}{2p}},$$

and bounding $\mathrm{meas}(\Omega^\sharp)$ as before yields

$$|\mathfrak{T}_{2,2}(\eta,e)| \lesssim h^{2k-1+\frac{p-2}{p}(k-1+\epsilon)}.$$

Collecting the above estimates and dropping higher-order terms, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}\|e\|_{L^2(\Omega)}^2 + n_{\mathrm{lim,ed}}(u_h;e,e) \lesssim h^{2k-1+\frac{p-2}{p}(k-1+\epsilon)} + h^{2k+1-\lambda\epsilon}.$$

For $k\geq 4$, we can take $\epsilon=0$ so that, with $p=6$, $2k-1+\frac{p-2}{p}(k-1+\epsilon) = 2k-1+\frac{2}{3}(k-1) \geq 2k+1$, yielding the desired estimate. In the case $k\leq 3$, we use a bootstrap argument. Suppose that the error estimate (3.8) holds with bound $h^{k+\rho_n}$. Then, the new bound for $\mathfrak{T}_{2,2}(\eta,e)$ is $h^{2k-1+\rho_n+\frac{p-2}{p}(k-1+\epsilon_n+\rho_n)}$, the bound on $\mathfrak{T}_{2,1}(\eta,e)$ remaining unchanged. The two terms are balanced by the choice

$$\epsilon_n = \frac{(3-k)p+2(k-1)}{(1+\lambda)p-2} - \frac{2(p-1)}{(1+\lambda)p-2}\rho_n$$

and this leads to the new error estimate with bound $h^{k+\rho_{n+1}}$ where $\rho_{n+1}=\frac{1}{2}(1-\lambda\epsilon_n)$. Hence, we obtain a recursive relation of the form $\rho_{n+1}=a+b\rho_n$. We want the fixed-point iteration to deliver a positive value. This, in turn, requires $a>0$. Moreover, the desired $\frac{1}{2}$ value is reached if $a+\frac{1}{2}b\geq\frac{1}{2}$. Taking $p=6$ yields

$$a = \frac{1}{2}\left(1-4\lambda\frac{4-k}{6\lambda+4}\right), \qquad b=\frac{5\lambda}{6\lambda+4}.$$

Condition $a>0$ is fulfilled for all λ if $k=3$, all $\lambda<2$ if $k=2$, and all $\lambda<\frac{2}{3}$ if $k=1$. The fixed-point is $\rho_\infty = \frac{2+\lambda(2k-5)}{4+\lambda}$, which is larger than $\frac{1}{2}$ only for $k=3$. This completes the proof. □

Remark 3.2. Sharper results can be derived by using the advective derivative in the weighting function. Following [17], the proof of the error estimate then hinges on discrete inf-sup stability, assuming additional regularity in time of the exact solution. As a result, the desired $h^{k+\frac{1}{2}}$ error estimate is achieved for any polynomial degree, any value for λ, and shape-regular meshes.

4. Numerical experiments. The objective of this section is to illustrate the efficiency of the weighted edge stabilization introduced and analyzed in §3.

4.1. One-dimensional tests. We report in this section one-dimensional tests over the periodic domain $\Omega := (0,1)$ using $\mathbb{P}_1$ finite elements. The time stepping is done with the SSP RK3 scheme with CFL $= 0.2$. The parameters for the first-order viscosity, entropy viscosity, and edge stabilization are $c_{\max} = \frac{1}{2}$, $c_{\rm ev} = \frac{1}{2}$, and $c_{\rm ed} = 0.05$, respectively.

4.1.1. Smooth transport. We consider the one-dimensional transport equation $\partial_t u + \partial_x u = 0$ with initial data $u_0(x) = \sin(2\pi x)$ to compare the convergence properties of the linear edge stabilization with those of the weighted edge stabilization. The computations are performed up to $t_{\rm F} = 1$ on various meshes. The meshes are non-uniform to avoid super-convergence effects. The size of each cell is chosen randomly with deformation factor equal to 3, that is, the size ratio between two neighboring cells is at most 3.

The L^1- and L^2-norms of the error at $t_F = 1$ are reported in Table 4.1 for seven meshes. This test shows that the convergence properties of the linear edge stabilization and those of the weighted edge stabilization are identical for smooth solutions.

(c) Edge stabilization

h	L^1-norm	rate	L^2 -norm	rate
1.000E-02	3.309E-04	–	4.095E-04	–
5.000E-03	7.845E-05	2.077	9.699E-05	2.078
2.500E-03	1.875E-05	2.065	2.342E-05	2.050
1.250E-03	4.630E-06	2.018	5.788E-06	2.017
6.250E-04	1.158E-06	1.999	1.439E-06	2.008
3.125E-04	2.874E-07	2.011	3.572E-07	2.010
1.563E-04	7.081E-08	2.021	8.796E-08	2.022

(d) Weighted edge stabilization

L^1-norm	rate	L^2-norm	rate
3.442E-04	–	5.233E-04	–
7.837E-05	2.135	1.068E-04	2.292
1.866E-05	2.070	2.448E-05	2.126
4.625E-06	2.013	5.954E-06	2.039
1.157E-06	1.999	1.467E-06	2.021
2.875E-07	2.009	3.625E-07	2.017
7.114E-08	2.015	8.939E-08	2.020

TABLE 4.1

Linear transport with smooth solution. L^1- and L^2-norms of error, $\mathbb{P}_1$ finite elements with edge stabilization (left) and weighted edge stabilization (right), $t_{\rm F} = 1$, non-uniform meshes.

4.1.2. Non-smooth transport. We now test the convergence properties of the entropy viscosity method with the weighted edge stabilization. We consider the one-dimensional transport equation with non-smooth initial data (2.11) and periodic boundary conditions. The computations are performed with $\mathbb{P}_1$ finite elements up to $t_{\rm F} = 1$ on various meshes. The L^1- and L^2-norms of the error are reported in Table 4.2 for seven uniform (left table) and seven non-uniform meshes (right table). The convergence orders in the L^1- and L^2-norm seem to be $\frac{3}{4}$ and $\frac{3}{8}$, respectively. Considering that the initial data is in $BV(\Omega)$ (almost $W^{1,1}(\Omega)$) and in $B^{\frac{1}{2}}_{\infty,2}(\Omega)$ (almost $H^{\frac{1}{2}}(\Omega) := W^{\frac{1}{2},2}(\Omega)$), these rates are compatible with the estimates $\frac{k+\frac{1}{2}}{k+1}$ and $\frac{1}{2}\frac{k+\frac{1}{2}}{k+1}$ obtained by interpolation using the rate $(k+\frac{1}{2})$ for smooth solutions.

We now verify that the weighted edge stabilization does preserve the weakened maximum principle (see (2.8)) of the entropy viscosity, contrary to the linear (unweighted) edge stabilization, see §2.2.2. We show in Table 4.3 the indicators $e_{\rm Max}$ and $e_{\rm Min}$ defined by (2.12) for seven uniform and non-uniform meshes. By comparing Table 2.1 with Table 4.3, we observe that the convergence rate on $e_{\rm Max}$ and $e_{\rm Min}$ for the entropy viscosity solution with weighted edge stabilization is larger than that for the entropy viscosity solution alone, so that the entropy viscosity and the weighted edge stabilization are now working together instead of antagonizing each other.

4.1.3. One-dimensional non-convex conservation equation. We finish this series of one-dimensional tests by testing again the nonlinear scalar conservation equation with non-convex flux

(a) Uniform meshes

h	L^1-norm	rate	L^2-norm	rate
1.000E-02	4.694E-02	–	1.194E-01	–
5.000E-03	2.720E-02	0.787	9.094E-02	0.393
2.500E-03	1.590E-02	0.774	6.965E-02	0.385
1.250E-03	9.342E-03	0.768	5.350E-02	0.381
6.250E-04	5.504E-03	0.763	4.115E-02	0.379
3.125E-04	3.254E-03	0.758	3.168E-02	0.378
1.563E-04	1.930E-03	0.754	2.439E-02	0.377

(b) Non-uniform meshes

L^1-norm	rate	L^2-norm	rate
4.897E-02	–	1.219E-01	–
2.873E-02	0.769	9.320E-02	0.388
1.685E-02	0.770	7.132E-02	0.386
1.002E-02	0.750	5.511E-02	0.372
5.954E-03	0.751	4.240E-02	0.378
3.552E-03	0.745	3.269E-02	0.375
2.115E-03	0.748	2.522E-02	0.374

TABLE 4.2

Linear transport with non-smooth solution. Entropy viscosity method with the weighted edge stabilization. L^1- and L^2-norms of error on uniform meshes (left) and non-uniform meshes (right), $\mathbb{P}_1$ finite elements, $t_F = 1$.

(a) Uniform meshes

h	e_{Min}	rate	e_{Max}	rate
1.000E-02	6.498E-03	–	6.515E-03	–
5.000E-03	4.470E-03	0.540	4.470E-03	0.544
2.500E-03	2.966E-03	0.592	2.966E-03	0.592
1.250E-03	2.021E-03	0.553	2.021E-03	0.553
6.250E-04	1.345E-03	0.588	1.345E-03	0.588
3.125E-04	8.992E-04	0.580	8.992E-04	0.580
1.563E-04	6.091E-04	0.562	6.091E-04	0.562

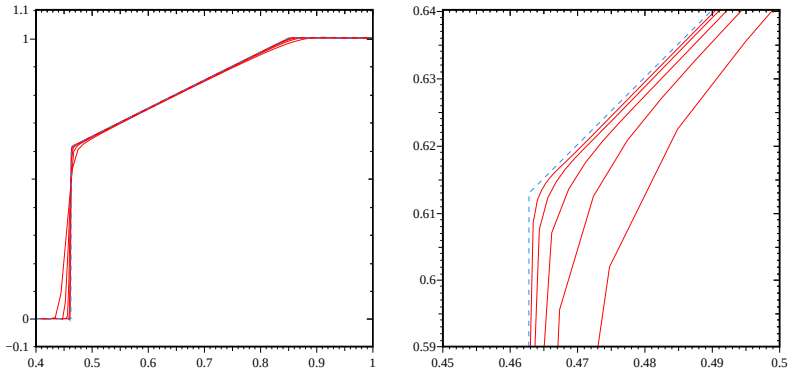
(b) Non-uniform meshes

e_{Min}	rate	e_{Max}	rate
6.655E-03	–	6.354E-03	–
4.452E-03	0.580	4.462E-03	0.510
3.171E-03	0.490	3.153E-03	0.501
2.124E-03	0.578	2.177E-03	0.534
1.474E-03	0.527	1.473E-03	0.564
1.013E-03	0.541	1.022E-03	0.527
7.544E-04	0.425	7.189E-04	0.508

TABLE 4.3

Weakened maximum principle for linear transport with non-smooth solution on uniform meshes (left) and non-uniform meshes (right), $\mathbb{P}_1$ finite elements, $t_F = 1$.

and initial data specified by (2.14). The computations are performed on five uniform meshes composed of 100, 200, 400, 800, and 1600 cells using the entropy viscosity method with weighted edge stabilization. The graphs of the solutions obtained at $t_F = 1$ are shown in Figure 4.1. We observe that, contrary to what is shown in Figure 2.2(b), the method now converges to the correct entropic solution thereby confirming again that weighting the edge stabilization is a good idea.



(a) Ent. visc. + weighted ed. st. (b) Ent. visc. + weighted ed. st. (zoom)

FIGURE 4.1. *Nonconvex flux. Results at $t_F = 1$ with entropy viscosity with weighted edge stabilization on five uniform meshes composed of* 100, 200, 400, 800, *and* 1600 *cells.*

4.2. Two-dimensional tests. We illustrate the method on two-dimensional problems.

4.2.1. Smooth transport. Let us consider the domain $\Omega = \{\boldsymbol{x} \in \mathbb{R}^2, \|\boldsymbol{x}\| < 1\}$, the rotating velocity field $\boldsymbol{\beta}(\boldsymbol{x}) = 2\pi(-y, x)$, where $\boldsymbol{x} = (x, y)$, and the two-dimensional transport problem

$$\partial_t u + \nabla\cdot(\boldsymbol{\beta} u) = 0, \qquad u(\boldsymbol{x}, 0) = \frac{1}{2}\left(1 - \tanh\left(\frac{(\boldsymbol{x} - \boldsymbol{r}_0)^2}{a^2} - 1\right)\right), \tag{4.1}$$

with $a = 0.3$ and $\boldsymbol{r}_0 = (0.4, 0)$. Note that $\nabla\cdot\boldsymbol{\beta} = 0$ and $\boldsymbol{\beta}\cdot\boldsymbol{n}|_{\partial\Omega} = 0$.

We perform tests on triangular isoparametric finite elements. The meshes are quasi-uniform and based on a Delaunay triangulation technique [28]. The time stepping uses the RK4 scheme and the CFL used in all the tests is CFL=0.25. The parameters for the first-order viscosity, entropy viscosity, and edge stabilization are $c_{\max} = \frac{1}{4k}$, $c_{\rm ev} = 0.1$, and $c_{\rm ed} = 0.025$, respectively.

Convergence tests on the above smooth problem, not reported here, have shown that, when used alone, the weighted edge stabilization performs as well as the original edge stabilization technique both for $\mathbb{P}_1$ and $\mathbb{P}_2$ finite elements. When combined with the weighted edge stabilization, the converge rate of the entropy viscosity method with $\mathbb{P}_1$ finite elements, which is already second-order in the L^2-norm, is not increased. The story is a little different for higher-order finite elements. The entropy viscosity alone has been observed to be second-order for $\mathbb{P}_1$ finite elements and to be of order $(k + \epsilon(k))$ for higher-order polynomial degrees; see [21]. Augmenting the entropy viscosity method with weighted edge stabilization improves the convergence order. We show in Table 4.4 convergence tests on the above smooth transport problem (4.1) with $\mathbb{P}_2$ finite elements. We compare the convergence rate of the entropy viscosity method and the entropy viscosity method with weighted edge stabilization. Adding the weighted edge stabilization clearly improves the convergence rate of the entropy viscosity method. Some super-convergence effect is observed (the errors are estimated by using a Gaussian quadrature exact on $\mathbb{P}_5$ polynomials).

(a) Entropy Viscosity

h	L_1	rate	L_2	rate
2.00E-01	1.102E-01	–	8.410E-02	–
1.00E-01	2.317E-02	2.251	1.700E-02	2.306
5.00E-02	3.659E-03	2.663	2.268E-03	2.906
2.50E-02	8.099E-04	2.176	4.579E-04	2.308
1.25E-02	2.142E-04	1.919	1.173E-04	1.965
1.00E-02	1.337E-04	2.109	7.370E-05	2.083

(b) Ent. Visc. + Weighted Ed. St.

L_1	rate	L_2	rate
1.104E-01	–	8.851E-02	–
1.783E-02	2.630	1.675E-02	2.401
1.362E-03	3.710	1.304E-03	3.684
1.121E-04	3.603	1.137E-04	3.520
9.368E-06	3.581	1.019E-05	3.479
4.249E-06	3.543	4.733E-06	3.439

TABLE 4.4

Linear transport with smooth solution (4.1). *L^1- and L^2-norms of error with the entropy viscosity method (left) and the entropy viscosity with weighted edge stabilization (right), $\mathbb{P}_2$ finite elements, $t_{\rm F} = 1$.*

4.2.2. Non-smooth transport. We solve again the linear transport problem (4.1), but this time we consider the following non-smooth initial data $u(\boldsymbol{x}, 0) = 1$ if $\|\boldsymbol{x} - \boldsymbol{r}_0\| < a$ and $u(\boldsymbol{x}, 0) = 0$ otherwise. The meshes and the time stepping are the same as before.

We report in Table 4.5 convergence tests in the L^1- and L^2-norms for various meshes using the entropy viscosity method with weighted edge stabilization. The results for $\mathbb{P}_1$ elements are shown in the left table and those for $\mathbb{P}_2$ elements are shown in the right table. The convergence orders in the L^1- and L^2-norms seem to be $\frac{3}{4}$ and $\frac{3}{8}$ for $\mathbb{P}_1$ elements and $\frac{4}{5}$ and $\frac{2}{5}$ for $\mathbb{P}_2$ elements, respectively. As above, these rates are compatible with the estimates $\frac{k+\frac{1}{2}}{k+1}$ and $\frac{1}{2}\frac{k+\frac{1}{2}}{k+1}$ obtained by interpolation using the rate $(k + \frac{1}{2})$ for smooth solutions.

We now verify the weakened maximum principle by computing the indicators $e_{\rm Max}$ and $e_{\rm Min}$ defined by (2.12) for each mesh. The results are reported in Table 4.6; those for $\mathbb{P}_1$ finite elements

(a) Ent. Visc. + Weighted Ed. St. $\mathbb{P}_1$

h	L_1	rate	L_2	rate
2.00E-01	1.381E+00	–	7.374E-01	–
1.00E-01	9.170E-01	0.591	5.544E-01	0.411
5.00E-02	5.454E-01	0.750	4.172E-01	0.410
2.50E-02	3.236E-01	0.753	3.158E-01	0.402
1.25E-02	1.904E-01	0.765	2.411E-01	0.389
1.00E-02	1.607E-01	0.759	2.214E-01	0.383
6.25E-03	1.124E-01	0.760	1.851E-01	0.381

(b) Ent. Visc. + Weighted Ed. St. $\mathbb{P}_2$

L_1	rate	L_2	rate
7.464E-01	–	5.058E-01	–
4.562E-01	0.710	3.779E-01	0.420
2.605E-01	0.808	2.794E-01	0.436
1.490E-01	0.806	2.114E-01	0.402
8.594E-02	0.794	1.601E-01	0.401
7.211E-02	0.786	1.466E-01	0.394

TABLE 4.5

Linear transport with non-smooth solution. L^1- and L^2-norms of error with entropy viscosity with weighted edge stabilization for $\mathbb{P}_1$ finite elements (left) and $\mathbb{P}_2$ finite elements (right), at $t_{\mathrm{F}} = 1$.

are shown in the left table and those for $\mathbb{P}_2$ finite elements in the right table. We observe that both indicators e_{Max} and e_{Min} converge to zero with the meshsize thus indicating that the entropy viscosity with weighted edge stabilization satisfies a weakened maximum principle.

(a) Ent. Visc. + Weighted Ed. St. $\mathbb{P}_1$

h	e_{Min}	rate	e_{Max}	rate
2.00E-01	4.264E-02	–	5.354E-01	–
1.00E-01	5.337E-02	-0.324	2.037E-01	1.394
5.00E-02	5.313E-02	0.006	3.546E-02	2.522
2.50E-02	1.911E-02	1.475	1.283E-02	1.467
1.25E-02	6.362E-03	1.587	7.776E-03	0.722
1.00E-02	5.713E-03	0.482	6.798E-03	0.603
6.25E-03	4.885E-03	0.333	5.461E-03	0.466

(b) Ent. Visc. + Weighted Ed. St. $\mathbb{P}_2$

e_{Min}	rate	e_{Max}	rate
2.003E-02	–	1.626E-01	–
1.250E-02	0.680	1.015E-02	4.001
8.397E-03	0.574	7.904E-03	0.361
7.900E-03	0.088	6.943E-03	0.187
6.131E-03	0.366	5.953E-03	0.222
5.150E-03	0.782	5.211E-03	0.596

TABLE 4.6

Linear transport with non-smooth solution. Weakened maximum principle at $t_{\mathrm{F}} = 1$ for entropy viscosity with weighted edge stabilization with $\mathbb{P}_1$ finite elements (left) and $\mathbb{P}_2$ finite elements (right).

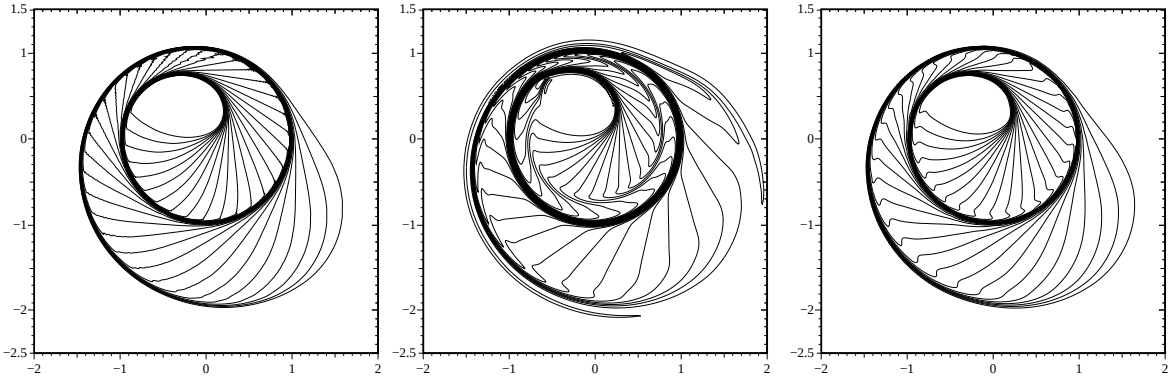
4.3. Non-convex conservation equation. Consider the following two-dimensional scalar conservation equation with non-convex fluxes:

$$\partial_t u + \nabla\cdot\boldsymbol{f}(u) = 0, \quad \boldsymbol{f}(u) = (\sin u, \cos u), \quad u(x,y,0) = \begin{cases} 3.5\pi, & x^2+y^2 < 1; \\ 0.25\pi, & \text{otherwise.} \end{cases} \tag{4.2}$$

The local velocity is $\boldsymbol{f}'(u) = (\cos u, -\sin u)$. This is a Cauchy problem in $\mathbb{R}^2$. This test was proposed in [27] and is challenging to many high-order numerical schemes because the solution has a two-dimensional composite wave structure; see [27] for details.

We perform computations on a mesh composed of triangular finite elements. The mesh family is quasi-uniform. The solution is computed at $t_F = 1$. The time stepping is done with the standard RK4 scheme with CFL = 0.025. Three computations are done with the entropy viscosity method (with coefficients $c_{\max} = \frac{1}{2}$, $c_{\mathrm{ev}} = 1$): one without edge stabilization, one with unweighted edge stabilization (with $c_{\mathrm{ed}} = 1$), and one with weighted edge stabilization (with $c_{\mathrm{ed}} = 1$ and $\lambda = 2$). The graph of the $\mathbb{P}_1$ solution without edge stabilization is shown in Figure 4.2(a). The solution shows the expected rotating composite wave structure composed of a shock and an expansion. This solution is an accurate approximation of the entropic solution. The graph of the solution with unweighted edge stabilization is shown in Figure 4.2(b). We observe that this solution exhibits a non-physical

layer between the shock and the expansion wave. We have verified by mesh refinement that this approximation does not converge to the entropic solution. Finally, the graph of the solution with weighted edge stabilization is shown in Figure 4.2(c). We observe that the unphysical layer has been pushed back to the shock, thereby recovering the correct physical behavior.



(a) Entropy visc., $c_{\max} = \frac{1}{2}$, $c_{\text{ev}} = 1$. (b) Ent. visc. + edge stab., $c_{\max} = \frac{1}{2}$, $c_{\text{ev}} = 1$. (c) Ent. visc. + weighted edge stab., $c_{\max} = \frac{1}{2}$, $c_{\text{ev}} = 1$.

FIGURE 4.2. *Two-dimensional conservation equation with non-convex flux,* $t_{\text{F}} = 1$. *Entropy viscosity solution without edge stabilization (left), with unweighted edge stabilization (center), and with weighted edge stabilization (right) on a uniform mesh composed of* 118850 $\mathbb{P}_1$ *nodes,* $CFL = 0.025$.

5. Conclusions. To conclude we would like to offer some thoughts on the relative importance of nonlinear viscosities (also called shock capturing in the literature) and linear stabilization techniques. Substantial efforts have been devoted to the construction of linear stabilization techniques (GaLS, RFB, SUPG, VMS, Streamline Diffusion, Subgrid Viscosity, Edge Stabilization, local projection, etc.), and it is often implicitly understood that linear stabilization is the workhorse whereas shock capturing is only meant to remove remaining spurious oscillations. As a result, the amount of research dedicated to constructing and analyzing shock capturing techniques is comparatively smaller than that dedicated to linear stabilization. As shown in the present paper, linear stabilization techniques can promote the Gibbs phenomenon and can transform a converging method into a non-converging one. We think that nonlinear stabilization methods (shock capturing, nonlinear viscosities, etc.) deserve more attention. These methods should be considered the workhorses that kill the Gibbs phenomenon and ensure convergence to the physically-correct weak solution; linear stabilization techniques should be seen as auxiliary tools whose job is to improve the convergence of an already converging nonlinear stabilization method. We refer to [5] for a recent step in this direction.

Appendix A. Entropy viscosity. The purpose of this appendix is to briefly recall the principles of the so-called entropy viscosity method. The specificities of the method are unrelated to the content of this paper, but we nevertheless describe how it is implemented for the sake of completeness. Full descriptions of the technique can be found in [20, 21]. The time approximation of (2.6) is done using an explicit Runge-Kutta method defined by a Butcher tableau

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & & \ddots & & \\ c_r & a_{r1} & a_{r2} & \cdots & a_{r,r-1} & \\ \hline & b_1 & b_2 & \cdots & b_{r-1} & b_r \end{array} \tag{A.1}$$

The algorithm is initialized by setting $\nu_h^0 = 0$ and $u_h^0 = u_{0,h}$. The viscosity field is piecewise constant over the mesh $\mathcal{K}_h$. Assume that the current time is t_n and let u_h^n be the approximation of $u_h(t_n)$. The next time increment is defined by

$$\Delta t_n = \mathrm{CFL} \min_{K\in\mathcal{K}_h} \frac{h_K}{|\boldsymbol{f}'(u_h^n|_K)|}. \tag{A.2}$$

The approximation u_h^{n+1} and the viscosity field ν_h^{n+1} are evaluated at time $t_{n+1} = t_n + \Delta t_n$ as

$$u_h^{n+1} = u_h^n - \Delta t_n(b_1k_1 + b_2k_2 + \cdots + b_rk_r), \qquad \nu_h^{n+1} = \nu_r, \tag{A.3}$$

where the functions $\{k_l\}_{1\le l\le r}$ and the viscosity fields $\{\nu_l\}_{1\le l\le r}$ are recursively computed as follows:

$$\left\{\begin{array}{l} w_1 = u_h^n, \\ \nu_1 = \nu_h^n, \quad \displaystyle\int_\Omega k_1 v\,\mathrm{d}\Omega = \int_\Omega v\nabla\cdot\boldsymbol{f}(w_1) + b_{\mathrm{ev}}(\nu_1, w_1, v), \quad \forall v\in X_h \\ w_2 = u_h^n - \Delta t_n a_{21}k_1, \\ \nu_2 = H_2(w_1, w_2), \quad \displaystyle\int_\Omega k_2 v\,\mathrm{d}\Omega = \int_\Omega v\nabla\cdot\boldsymbol{f}(w_2) + b_{\mathrm{ev}}(\nu_2, w_2, v), \quad \forall v\in X_h \\ \vdots \\ w_r = u_h^n - \Delta t_n(a_{r1}k_1 + a_{r2}k_2 + \cdots + a_{r,r-1}k_{r-1}) \\ \nu_r = H_r(w_1, w_2, \ldots, w_r), \displaystyle\int_\Omega k_r v\,\mathrm{d}\Omega = \int_\Omega v\nabla\cdot\boldsymbol{f}(w_r) + b_{\mathrm{ev}}(\nu_r, w_r, v), \quad \forall v\in X_h, \end{array}\right. \tag{A.4}$$

where $b_{\mathrm{ev}}(\nu, w, v) := \sum_{K\in\mathcal{K}_h} \nu|_K \int_K \nabla w\cdot\nabla v\,\mathrm{d}K$.

The functions $H_2, \ldots, H_r$ are piecewise constant over the mesh $\mathcal{K}_h$ and are defined as follows. First, we define the so-called entropy functional $E \in \mathcal{C}^1(\mathbb{R};\mathbb{R})$. Second, we evaluate the entropy residual associated with two given states $c_h, d_h \in X_h$ as follows. Let δ be the time increment separating the states c_h and d_h. The entropy residual in the cell K at $\boldsymbol{x}\in K$ is defined as

$$R_K(c_h, d_h, \delta)(\boldsymbol{x}) = \frac{E(d_h(\boldsymbol{x})) - E(c_h(\boldsymbol{x}))}{\delta} + \boldsymbol{f}'(d_h(\boldsymbol{x}))\cdot\nabla(E(d_h(\boldsymbol{x}))). \tag{A.5}$$

We now define the entropy residual associated with the mesh interfaces. Let $F\in\mathcal{F}_h^i$. Then, for all $\boldsymbol{x}\in F$, we set

$$J_F(d_h)(\boldsymbol{x}) = (\boldsymbol{f}'(d_h(\boldsymbol{x}))\cdot\boldsymbol{n}(\boldsymbol{x}))[\![\partial_n E(d_h(\boldsymbol{x}))]\!]. \tag{A.6}$$

The total entropy residual in the cell K is then defined to be

$$D_K(c_h, d_h, \delta) = \max(\|R_K(c_h, d_h, \delta)\|_{L^\infty(K)}, \max_{F\in\partial K\cap\mathcal{F}_h^i} \|J_F(d_h)\|_{L^\infty(F)}). \tag{A.7}$$

The maximum local wave speed in the cell K is estimated by

$$\beta_K(d_h) = |\max_{\boldsymbol{x}\in K} \boldsymbol{f}'(d_h(\boldsymbol{x}))|. \tag{A.8}$$

Then, the value over K of the viscosity function H_l, $2\le l\le r$, is defined as follows:

$$H_l(w_1, \ldots, w_l)|_K = \min(c_{\max}\beta_K(w_l)h_K, c_{\mathrm{ev}}h_K^2 D_K(w_1, w_l, c_l\Delta t_n)) \tag{A.9}$$

Note that each step of (A.4) requires solving a mass matrix linear system. We stress again that we do not recommend to use the lumped mass matrix instead of the consistent mass matrix since mass lumping induces large dispersion errors.

REFERENCES

[1] Y. Achdou, C. Bernardi, and F. Coquel. A priori and a posteriori analysis of finite volume discretizations of Darcy's equations. *Numer. Math.*, 96(1):17–42, 2003.

[2] M. Boman. Estimates for the L_2-projection onto continuous finite element spaces in a weighted L_p-norm. *BIT*, 46(2):249–260, 2006.

[3] S. Brenner and R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, 1994.

[4] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.

[5] E. Burman. On nonlinear artificial viscosity, discrete maximum principle and hyperbolic conservation laws. *BIT*, 47(4):715–733, 2007.

[6] E. Burman and A. Ern. Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.*, 191(35):3833–3855, 2002.

[7] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.*, 76(259):1119–1140, 2007.

[8] E. Burman, A. Ern, and M. A. Fernández. Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems. *SIAM J. Numer. Anal.*, 48(6):2019–2042, 2010.

[9] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(15-16):1437–1453, 2004.

[10] R. Codina. A discontinuity-capturing crosswind-dissipation for the finite element solution of the convection-diffusion equation. *Comput. Methods Appl. Mech. Engrg.*, 110(3-4):325–342, 1993.

[11] M. Crouzeix and V. Thomée. The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.

[12] D. A. Di Pietro and A. Ern. *Mathematical Aspects of Discontinuous Galerkin Methods*, volume 69 of *Mathématiques et Applications*. Springer, Heidelberg, 2012.

[13] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[14] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 44(2):753–778, 2006.

[15] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112 (electronic), 2001.

[16] J.-L. Guermond. Stabilization of Galerkin approximations of transport equations by subgrid modeling. *M2AN Math. Model. Numer. Anal.*, 33(6):1293–1316, 1999.

[17] J.-L. Guermond. Subgrid stabilization of Galerkin approximations of linear contraction semi-groups of class C^0 in Hilbert spaces. *Numer. Methods Partial Differential Equations*, 17:1–25, 2001.

[18] J.-L. Guermond. On the use of the notion of suitable weak solutions in CFD. *Int. J. Nmer. Methods Fluids*, 57:1153–1170, 2008.

[19] J.-L. Guermond, A. Marra, and L. Quartapelle. Subgrid stabilized projection method for 2D unsteady flows at high Reynolds number. *Computer Methods in Applied Mechanics and Engineering*, 195:5857–5876, 2006.

[20] J.-L. Guermond and R. Pasquetti. Entropy-based nonlinear viscosity for Fourier approximations of conservation laws. *C. R. Math. Acad. Sci. Paris*, 346:801–806, 2008.

[21] J.-L. Guermond, R. Pasquetti, and B. Popov. Entropy viscosity method for nonlinear conservation laws. *Journal of Computational Physics*, 230:4248–4267, 2011.

[22] J. S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods*, volume 54 of *Texts in Applied Mathematics*. Springer, New York, 2008. Algorithms, analysis, and applications.

[23] T. Hughes and M. Mallet. A new finite element formulation for computational fluid dynamics. IV: A discontinuity-capturing operator for multidimensional advective-diffusive systems. *Comput. Methods Appl. Mech. Engrg.*, 58:329–336, 1986.

[24] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46(173):1–26, 1986.

[25] C. Johnson and A. Szepessy. On the convergence of a finite element method for a nonlinear hyperbolic conservation law. *Math. Comp.*, 49(180):427–444, 1987.

[26] O. A. Karakashian and F. Pascal. A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems. *SIAM J. Numer. Anal.*, 41(6):2374–2399, 2003.

[27] A. Kurganov, G. Petrova, and B. Popov. Adaptive semidiscrete central-upwind schemes for nonconvex hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 29(6):2381–2401 (electronic), 2007.

[28] S. Rebay. Efficient unstructured mesh generation by means of Delaunay triangulation and Bowyer-Watson algorithm. *J. Comput. Phys.*, 106:125–138, 1993.